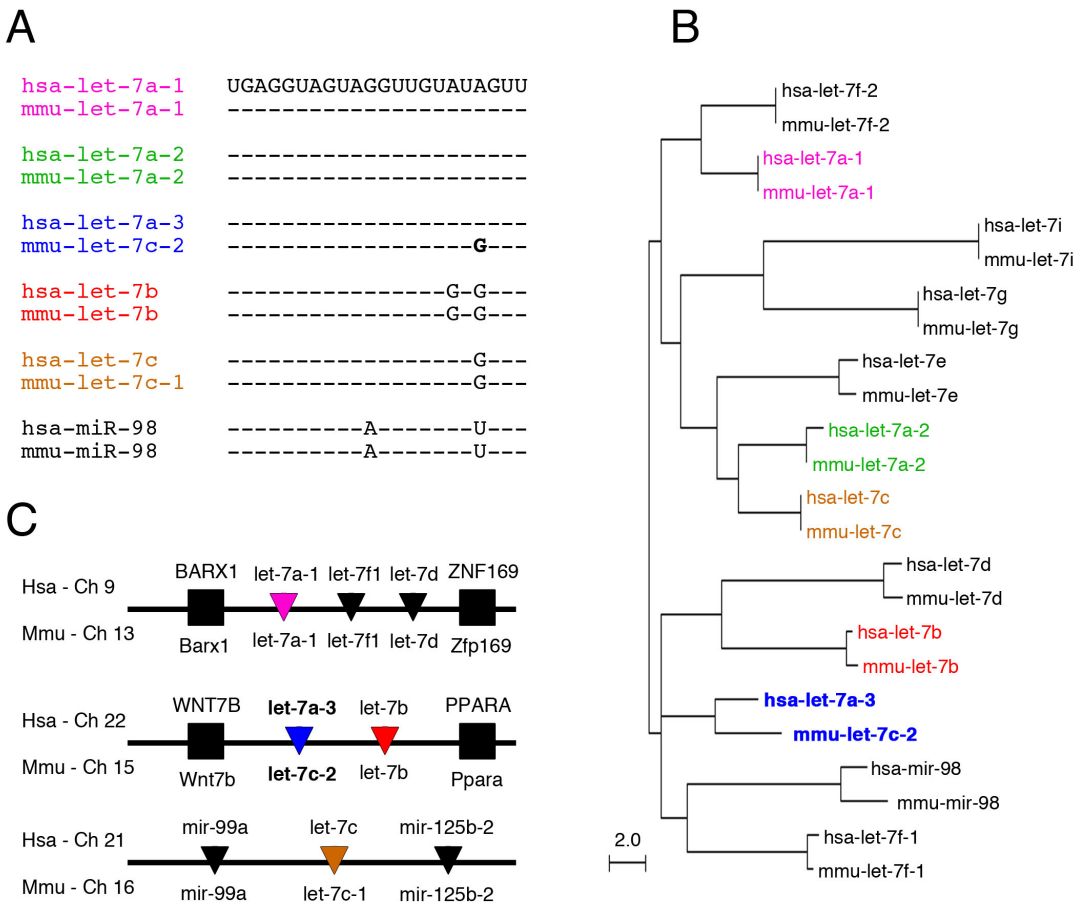
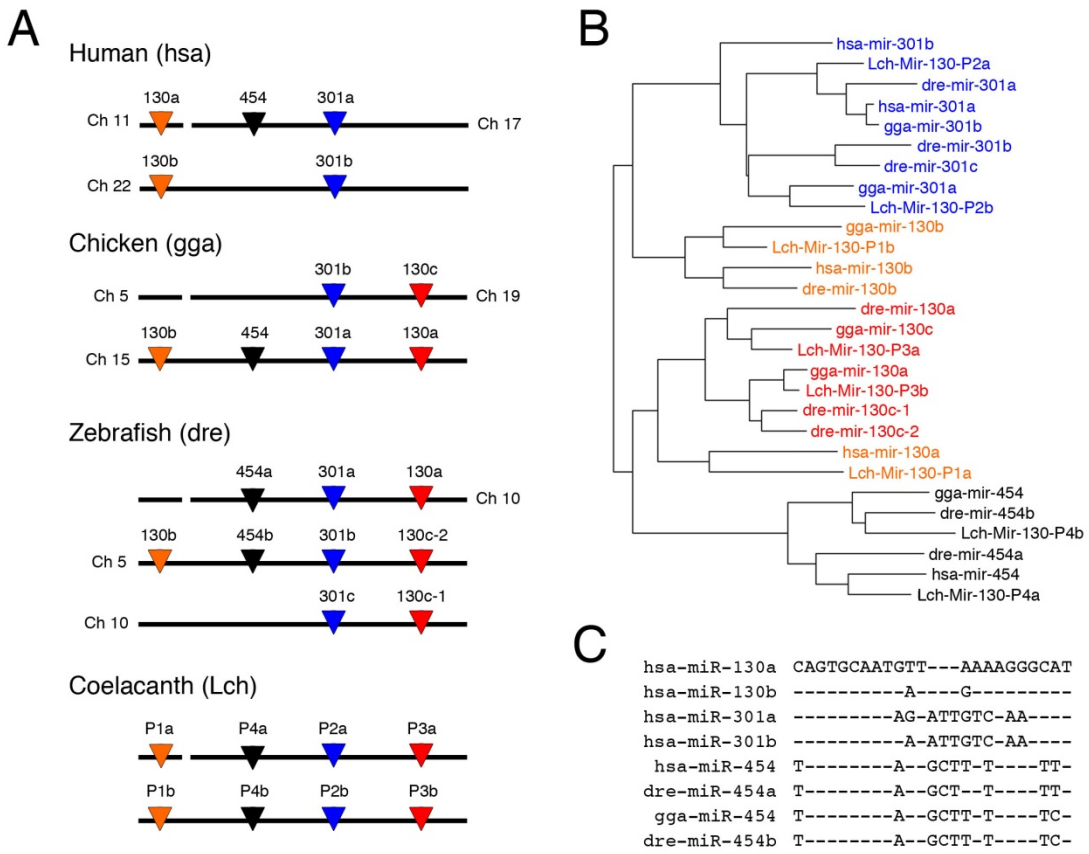


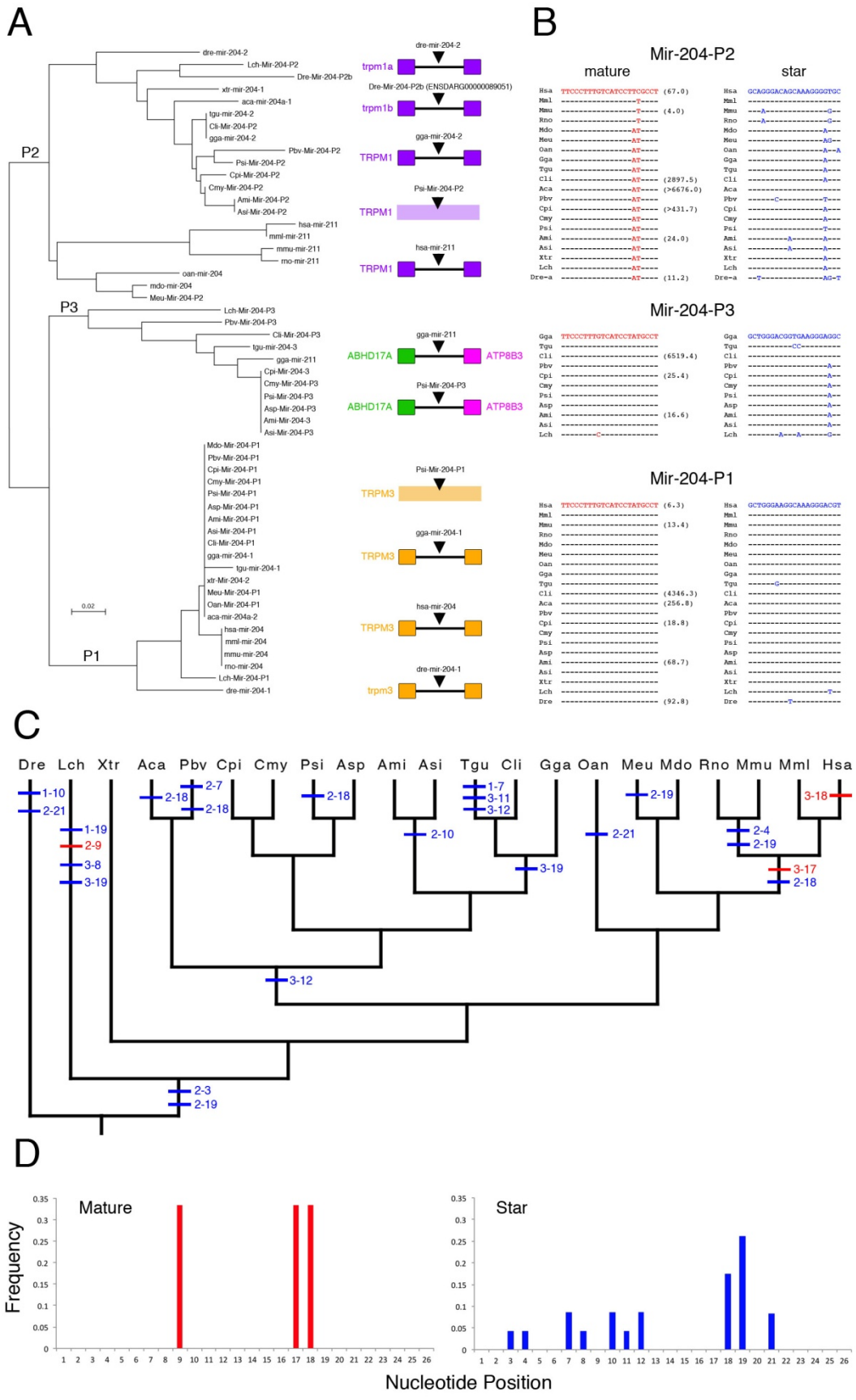
Supplemental Figure 1. Single nucleotide substitutions change the name of a sequence and hence obfuscate evolutionary understanding. **A.** Shown are alignments of six let-7 sequences from human and mouse, aligned to the human let-7a-1 sequence. Dashes indicate identity with this human sequence. Note that the human (hsa) let-7c sequence and the mouse (mmu) let-7c-1 sequence (orange) as well as the mmu-let-7c-2 sequence (blue) have a "G" in position 19 (bold). However, the mouse let-7c-2 gene phylogenetically groups with the human let-7a-3 sequence (**B**, blue) and sits in the same relative genomic position, between the let-7b gene (**C**, red) and the protein-coding gene *Wnt7b*. Hence, orthologous genes are given different names in these two closely related

taxa due to a single convergent nucleotide substitution in the mouse sequence relative to the human sequence.



Supplemental Figure 2. Differential loss can create confusion recognizing orthologues from paralogues. **A.** The mir-130 family primitively (at least for bony fish) consists of eight genes, four linked in a single cluster, three in another, and one isolated gene. This is inferred from the genomic organizations of these genes in human, chicken, zebrafish and coelacanth. All of these genes are present in the coelacanth, but differential loss of some of these genes in the human, chicken and zebrafish lineages results in incomplete clusters of genes, with the human mir-454 gene associated with the "a" cluster and the chicken mir-454 gene associated with the "b" cluster. Coelacanth genes are named according to the nomenclature proposed herein (see Fig. 4, Supp. Fig. 3, and Supp. File 3, as well as the text). **B.** A mid-point rooted phylogenetic (distance) analysis supporting the scenario outlined in **A**. Because of the assumed homology between the two mir-454 genes, the

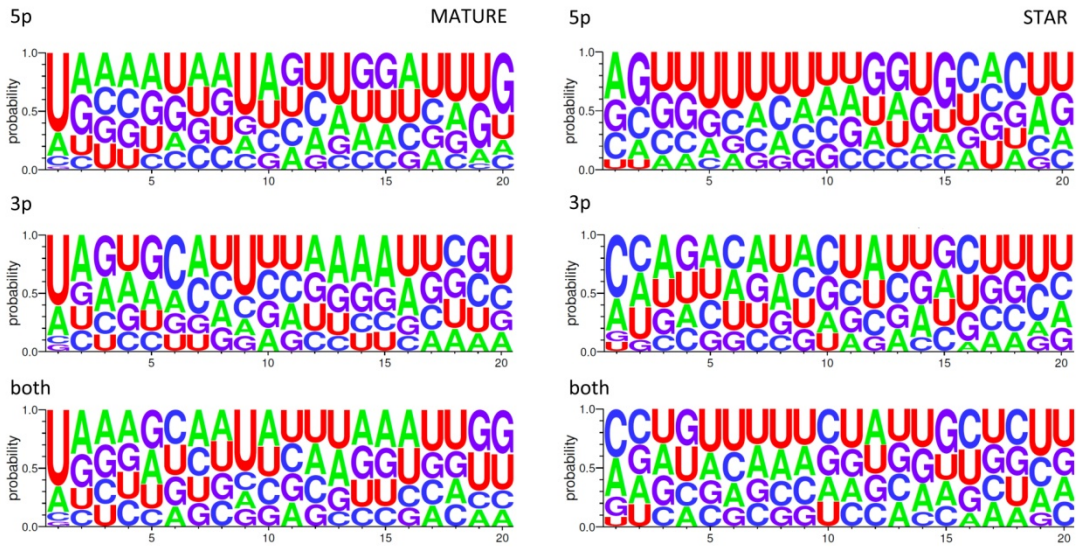
chicken “b” cluster is given the same names as the genes found in the human “a” cluster, but these are all paralogues of one another, not orthologues, and thus the human mir-454 gene is paralogous, not orthologous, to the chicken mir-454 gene. **C.** Alignment of the mir-130 mature sequences showing the identity of the seed region between the mir-130, mir-301 and mir-454 sequences, in addition to other regions of similarity among the three sets of genes.


A
B
C
D
P2
P3
P1
Mir-204-P2
mature
star
Mir-204-P3
Mir-204-P1
Nucleotide Position
Frequency
Mature
Star

Supplemental Figure 3. The nomenclature proposal and methods employed within the manuscript. **A.** The proposed nomenclature system for the mir-204 family, highlighting the advantages of the new system and the disadvantages of the old. Shown on the left is a phylogenetic tree of all of the mir-204 relatives in 21 osteichthyian taxa. Genes already deposited in miRBase are named according to the miRBase entry (e.g., hsa-mir-204); those that are not yet released are given the proposed name herein (e.g., Mdo-Mir-204-P1). Note that there are three clear Mir-204 paralogues, here named P1, P2 and P3. Mir-204-P1 is present in all 21 taxa, and in all taxa with a sequenced genome this gene is located in an intron of the TRPM3 gene (or associated with the gene where exon/intron boundaries are not defined, as in the turtle *Pelodiscus sinensis*) (orange). The Mir-204-P2 gene is also present in all 21 taxa, and is located in (or associated with) an intron in the TRPM1 gene (purple). Note though that there are two copies of this gene in zebrafish, and both are associated with paralogues of the trpm1 gene (top); one of these genes is novel in that no miRBase entry is recorded for this gene, but an identifier is present in Ensembl. The third Mir-204 paralogue – Mir-204-P3 – is absent in mammals, zebrafish, and *Xenopus*, but present in all of the other taxa and in both birds and turtles is linked to the ABHD17A (green) and ATB8B3 (magenta) genes. Thus, both the phylogenetic and the syntenic analyses support the proposed nomenclature. Importantly, the eutherian mir-211 sequence belongs to the Mir-204-P2 subgroup (which incudes the only Mir-204 gene in the platypus [oan-mir-204] and opossum [mdo-mir-204] currently deposited in miRBase), whereas the chicken mir-211 sequence belongs to the Mir-204-P3 subgroup. Furthermore, given that the platypus and opossum mir-204 sequences are members of the P2 subclade, there are likely genes in one or both taxa related to the P1 subgroup, and

indeed both of these genes are present in their respective genomes, but are not yet deposited in miRBase. Species abbreviations here and in all other figures and tables are as follows: Aca - *Anolis carolinensis* (lizard); Ami - *Alligator mississippiensis* (American alligator); Asi - *Alligator sinensis* (Chinese alligator); Asp - *Apalone spinifera* (spiny softshell turtle); Cli - *Columba livia* (pigeon); Cmy - *Chelonia mydas* (sea turtle); Cpi - *Chrysemys picta* (painted turtle); Dre - *Danio rerio* (zebrafish); Gga - *Gallus gallus* (chicken); Hsa - *Homo sapiens* (human); Lch - *Latimeria chalumnae* (coelacanth); Meu - *Macropus eugenii* (Tammar wallaby); Mdo - *Monodelphis domestica* (opossum); Mml – *Macaca mulatta* (rhesus macaque); Mmu - *Mus musculus* (mouse); Oan - *Ornithorhynchus anatinus* (platypus); Pbi - *Python bivittatus* (snake); Psi - *Pelodiscus sinensis* (Chinese softshell turtle); Rno – *Rattus norvegicus* (rat); Tgu - *Taenopygia guttata* (zebrafinch); Xtr - *Xenopus tropicalis* (frog). **B.** The procedure used to tabulate the nucleotide substitutions in pre-miRNA sequences. Shown are the alignments (using ClustalW, Macvector v. 10.0.2) of both the mature (red) and star (blue) reads from all Mir-204 genes from all 21 considered taxa. The numbers in parentheses to the right of the mature read indicated the fold difference between mature and star (from MirBase v. 21 and ref. 39). The dashes indicate identify between the species and human (or chicken for paralogue 3), whereas nucleotide substitutions relative to each of the human (or chicken) paralogues are shown as distinct nucleotides. **C.** A phylogenetic reconstruction of each of these nucleotide substitutions in the context of a known phylogenetic tree (39). The mutations to the mature are shown in red, whereas mutations to the star shown are shown in blue. The paralogue number and nucleotide position of each mutation for both the mature and star were calculated using MacClade (v. 4.08). For changes at the root of the

tree, paralogues and/or outgroups comparisons were used. For example, in Mir-204-P1, there is a change in the star at position 10 such that zebrafish (Dre) possesses a T whereas all of the other taxa possess a G. Because the G is present in this position in the other two paralogues (P2, and P3) the G is reconstructed as the primitive nucleotide for this position rather than the T. **C.** These mutations are then plotted by position as a frequency histogram for both the mature (red) and star (blue) sequences. This was done for each of the 234 genes reconstructed as present in the last common ancestor of tetrapods (see Supp. File 5) and the results are presented in Fig. 6.



Supplemental Figure 4. Nucleotide profiles of mature and star sequences of 234 genes present in the last common ancestor of tetrapods separated into 5p versus 3p arms. Note that although there are a few interesting differences between the two arms, the major biases shown in Figure 6 are seen in both the 5p and 3p arms and hence these biases are arm independent.

Supplemental File 1. Measurements on read length, complementarity, loop length, and time of acquisition for all accepted miRNA genes in five taxa, human, zebrafish, nudibranch, fruit fly and nematode. These are the data for what is presented in Table 1, Figure 3A, and Figure 5. The mature and star sequences for human from genes that were acquired by the time of the last common ancestor of tetrapods were used to generate the sequence logos in Fig. 6D.

Supplemental File 2. Our evaluation of all metazoan microRNAs deposited at miRBase as of version 21. These are the data for Figure 3B.

Supplemental File 3. The proposed nomenclature system for all multi-gene families for six vertebrate taxa, human, mouse, chicken, anolis lizard, frog and zebrafish.

Supplemental File 4. The gain and loss of miRNAs for every node given in Figure 5.

Supplemental File 5. The data for the mutational profiles of all 234 miRNA genes reconstructed as present in the last common ancestor of tetrapods. These are the data for the mutational profiles presented in Fig. 6A and C, and the rates estimates presented in Fig. 6B and Table 3.

Supplemental File 6. miRNAs deposited at MirBaseDB.org that either differ in sequence (red) and/or genome coordinates (yellow) from what is deposited at miRBase. Concordant entries are shown in green.